Metrika (1999) 50: 163–172

Metrika


# The mixture index of fit and minimax regression

**Tamás Rudas**

Department of Statistics, Eötvös Loránd University and TÁRKI, Pollack M. tér 10., H-1088 Budapest, Hungary (e-mail: rudas@tarki.hu)

Received: September 1997

**Abstract.** A measure of the fit of a statistical model can be obtained by estimating the relative size of the largest fraction of the population where a distribution belonging to the model may be valid. This is the mixture index of fit that was suggested for models for contingency tables by Rudas, Clogg, Lindsay (1994) and it is extended here for models involving continuous observations. In particular, the approach is applied to regression models with normal and uniform error structures. Best fit, as measured by the mixture index of fit, is obtained with minimax estimation of the regression parameters. Therefore, whenever minimax estimation is used for these problems, the mixture index of fit provides a natural approach for measuring model fit and for variable selection.

**Key words:** Minimax estimation, minimum distance estimation, mixture index of fit, regression with uniform error, variable selection

## 1 Introduction

When testing the fit of a statistical hypothesis, the achieved probability level of the data is often interpreted as a measure of support provided by the data for the hypothesis. This practice is not satisfactory for logical and practical reasons (see Schervish, 1996). The latter include that very often only an approximation of the distribution of the test statistic is available but the error of this approximation is not known. A typical problem where measuring the evidence for or against a hypothesis is crucial is that of model selection.

The difficulties associated with the goodness-of-fit interpretation of $p$-values are very severe in the context of contingency tables. Here, most of the testing relies on statistics having an asymptotic chi-squared distribution. In the case of small samples, the reference distribution is not valid, and for larger samples the researchers very often experience that, disappointingly, the desired simple models have to be rejected (because for samples yielding the same observed distribution, the values of test statistic are proportional to the sample size). To handle these problems, Rudas, Clogg, Lindsay (1994) suggested a general framework to measure model fit for contingency tables. The assumption that a probability distribution belonging to the model of interest may describe the entire population is dropped. The true distribution is supposed to be a mixture of two distributions, one is from the model and the other one is completely unrestricted. In this mixture, the weight of the distribution from the model is $1 - \pi$ and the weight of the other, unrestricted, distribution is $\pi$. Note that this assumption is nonrestrictive, as it always holds true for some value of $\pi$. The maximal value of $1 - \pi$ with which such a representation is possible, is a measure of the goodness-of-fit of the model. Here, the mixture index of fit is defined as a population parameter. When data are available, it can be estimated by the method of maximum likelihood, and a confidence interval can be constructed for its true value. For details see Rudas et al. (1994). Properties and applications of the mixture index of fit were explored in a series of papers, including Clogg, Rudas, Xi (1995), Xi (1996), Clogg, Rudas, Matthews (1998), Rudas, Zwick (1997).

There are many other areas in statistics, where measuring model fit or comparing the fits of several models is relevant, most notably regression analysis. The goal of the present paper is to extend the mixture index of fit to this problem, including stepwise variable selection.

In Section 2, the mixture index of fit will be defined for any hypothesis consisting of a family of dominated distributions. The related minimum distance estimation (minimum mixture estimation) problem will be shown to lead to minimax estimation. In Section 3, the theory developed in Section 2 will be applied to the problem of regression analysis with normal error structure. In Section 4, our approach will be applied to regression analysis with uniform error. The application of the mixture index of fit in this case is particularly useful, because when normal error distribution can be assumed, there is a well-developed statistical theory available to measure model fit, and partly because minimax regression was shown to perform better with short tailed error distributions than least squares regression (see Fabian, 1988, Narula, Wellington, 1985, Sklar, Armstrong, 1984). Finally, it will be illustrated in Section 5, by the analysis of a set of petrol refinery data which was analyzed earlier by Wood (1973) and Narula, Wellington (1985), that the mixture index of fit has a straightforward interpretation with a natural calibration that can be used for model selection.

## 2 The mixture index of fit

Let us consider a sample space and the collection $F$ of the densities of all probability distributions on the sample space which are dominated by a fixed measure. Let $G \quad F$ be the family of primary interest. This family is determined by the observational procedure of the data or by subject matter considerations, and should be rich enough to contain all densities of interest. The family $G$ is defined by the assumptions which are not tested in the statistical analysis. Particular choices are illustrated in Sections 3 and 4. A subset $H \quad G$ is a statistical model. The density of the distribution generating the observations is assumed to belong to $G$, and it is tested whether or not it is reasonable to believe that it also belongs to $H$.

All the assertions which follow for densities are true with probability 1, but reference to this fact will be suppressed. Also, the arguments of the density functions will not be shown.

For a fixed density, $g \in G$, the mixture index of fit $\pi^*(g, H)$ is the value of the following deviation

$$\pi^*(g, H) = \inf(\pi : g = (1 - \pi)h + \pi f, h \in H, f \in F, 0 \le \pi \le 1).$$

The interpretation associated with $\pi^*(g, H)$ is that the density $g$ is represented as a mixture of two densities, $h$ from the model $H$ and another unrestricted one, $f$ from $F$. The maximum possible weight a density from $H$ may have in such a representation, $1 - \pi^*$, is a measure of the closeness of $g$ to $H$, and the minimum possible weight the other component needs to have, $\pi^*$, is a measure of deviation of $g$ from $H$.

If $g$ is the density of the distribution of the population, then the framework here replaces the usual hypothesis that

$$g \in H \tag{1}$$

by the assumption that

$$g = (1 - \pi)h + \pi f, h \in H. \tag{2}$$

While the hypothesis (1) may or may not hold true, the assumption in (2) is always valid for some value of $\pi$ between 0 and 1, because for $\pi = 1$, (2) is not restrictive at all. Therefore, (2) is a framework which, depending on the data available, may lead to a valid analysis, even in cases when model (1) is not true and an analysis based on it may not be relevant at all. For a discussion of how the framework in (2) leads to a new kind of residual analysis, see Clogg et al. (1995, 1997).

The value of $\pi^*$ is derived from the always valid framework (2); $\pi^*$ is a measure of badness-of-fit and $1 - \pi^*$ is a measure of goodness-of-fit.

A related deviation can be defined for any two densities $g, h \in F$ as follows

$$\pi^*(g, h) = \inf(\pi : g = (1 - \pi)h + \pi f, f \in F, 0 \le \pi \le 1). \tag{3}$$

*Theorem.*

$$\pi^*(g, H) = 1 - \sup_{h \in H} \inf_{\operatorname{supp} h}(g/h),$$

where $\operatorname{supp} h = \{x : h(x) > 0\}$ is the support of $h$ and the infimum of $g/h$ is taken on $\operatorname{supp} h$.

*Proof.* In view of (3), it is sufficient to show that

$$\pi^*(g, h) = 1 - \inf_{\operatorname{supp} h}(g/h). \tag{4}$$

First note that $\pi^*(g, h) = 0$ if and only if $g = h$, and in that case, (4) holds true.

If $g \neq h$ then $\inf_{\mathrm{supp}\, h}(g/h) \quad 1$, otherwise $g \quad h$ would follow, which implies that $g = h$. Therefore,

$$f^* = \frac{1}{1 - \inf_{\mathrm{supp}\, h}(g/h)}(g - (\inf_{\mathrm{supp}\, h}(g/h))h)$$

is non negative, measurable, and its integral is 1, that is, $f^* \in F$. Thus,

$$g = \inf_{\mathrm{supp}\, h}(g/h)h + (1 - \inf_{\mathrm{supp}\, h}(g/h))f^*$$

is a mixture representation of the form (2) and, therefore

$$\pi^*(g,h) \quad 1 - \inf_{\mathrm{supp}\, h}(g/h). \tag{5}$$

On the other hand, for any representation of the form (2),

$$g/h = (1 - \pi) + \pi f/h \quad 1 - \pi$$

on supp $h$. Thus, $\inf_{\mathrm{supp}\, h}(g/h) \quad 1 - \pi$ that is, $\pi \quad 1 - \inf_{\mathrm{supp}\, h}(g/h)$, implying that $\pi^*(g,h) \quad 1 - \inf_{\mathrm{supp}\, h}(g/h)$ . This, together with (5) completes the proof.

Note that the proof also demonstrates that a mixture representation yielding the mixing weight is $\pi^*(g,h)$ is always possible in (3). The Theorem implies that

$$\pi^*(g,h) = \inf_{h \in H} \sup_{\mathrm{supp}\, h} \frac{h - g}{h},$$

that is, $\pi^*$ is the optimal value of the criterion function of relative minimax estimation.

In statistical applications, the model $H$ is specified but the true density $g$ of the underlying population is not known. As $\pi^*(g,H)$ is a uniquely defined functional, an estimate of its value is $\pi^*(\tilde{g},H)$, where $\tilde{g}$ is an estimate of the density $g$, based on the available data.

If the fit of the model $H$ is measured by (an estimate of) the mixture index of fit $\pi^*(g,H)$, then it is natural to consider the density $h^* \in H$ for which

$$\pi^*(g,h^*) = \pi^*(g,H) \tag{6}$$

as the minimum mixture estimate of the density of the distribution in the fraction of the population where $H$ holds true.

By combining (4) and (6), one obtains that

$$\pi^*(g,H) = 1 - \inf_{\mathrm{supp}\, h^*}(g/h^*). \tag{7}$$

This formula can be used to compute the value of the mixture index of fit once the minimum mixture estimate is determined.

## 3 Minimum mixture estimation in regression with normal error

In this section, a simple regression model is considered and minimum mixture estimation is applied to estimate the parameters.

The observations to analyze are $(x, y)$ pairs, where $x$ is possibly vector valued and $y$ is considered to be a response variable to $x$. To build an appropriate theoretical framework, let the dominating measure be the Lebesgue measure on the $n+1$ dimensional Euclidean space, if $x$ is $n$ dimensional, and $F$ is the collection of all densities with respect to the Lebesgue measure. The primary family of interest, $G$, consist of joint densities of variables $X$ and $Y$, where the conditional distribution of $Y$ given that $X = x$ is normal, with expectation $\mu_x$ and standard deviation $\sigma$.

The variable $Y$ is observed at $k$ different fixed values of $X$, namely $x_1, \dots,$ $x_k$, i.e., $X$ is not assumed to be random. In this situation, estimation is usually based on the conditional densities, expressing the fact that fit is equally important for all values of the independent variable $X$ (see Neter, Kutner, Nachtsheim, Wasserman, 1996). One convenient way to define joint distributions of $X$ and $Y$ with the same estimates as the ones obtained from the conditional densities, is to assign uniform marginal distribution to $X$. Therefore, the densities in $G$ have the following form

$$g(x, y) = \frac{1}{k} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2} \frac{(y - \mu_x)^2}{\sigma^2}\right] \quad \text{if } x \in \{x_1, \dots, x_k\},$$

for every real $y$, for some $\mu_x$ and $\sigma$.

The model $H$, in addition, assumes that the conditional expectation of $Y$ given $X = x$ is a linear function of the condition, that is, $E(Y|X = x) = a + b'x$, with some unspecified $a$ and $b$, where $b$ is a vector of appropriate dimension. The densities in $H$ have the form

$$h(x, y) = \frac{1}{k} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2} \frac{(y - a + b'x)^2}{\sigma^2}\right] \quad \text{if } x \in \{x_1, \dots, x_k\}.$$

In order to test the additional assumption in $H$ by the mixture approach, a smoothed version of the data, that is, an initial estimate in $G$, $\tilde{g}$, is needed. This, as usual, is obtained by replacing $\mu_x$ and $\sigma$ by estimates $m_x$, the conditional mean at $X = x$, and $s$, the pooled estimate of the standard deviation, respectively. Then, $\tilde{g}$ is represented as a mixture, with one component from $H$ and the minimization of $\pi$ is carried out. In the present case, unless $\sigma < s$, $\pi$ in (2) is 1, and therefore, in order to minimize $\pi$, the $\sigma$ $s$ values in $H$ can be disregarded.

To obtain the minimum mixture estimate of $\tilde{g}$ in $H$ and to estimate the value of $\pi^*$, first the infimum of $\tilde{g}/h$ should be determined. Straightforward calculations show that for fixed $a, b, \sigma, x$ , where $x$ is one of the observed values, the value of $Y$ yielding the infimum is

$$y = \frac{s^2(a + b'x) - \sigma^2 m_x}{s^2 - \sigma^2}$$

and the value of the infimum is

$$\frac{\sigma}{s}\exp\left[\frac{-(a+b'x-m_x)^2}{2(s^2-\sigma^2)}\right]. \tag{8}$$

Next, the value of $X$ yielding the infimum should be selected. As the exponent in (8) is non-positive, the infimum of (8) occurs for the $x$ value, for which $(a+b'x-m_x)^2$ is maximal. With this choice, for fixed $a, b, \sigma$,

$$\inf_{\operatorname{supp} h}\frac{\tilde{g}}{h}=\frac{\sigma}{s}\exp\left[\frac{-\max_x(a+b'x-m_x)^2}{2(s^2-\sigma^2)}\right]. \tag{9}$$

To determine the supremum of this quantity in $H$, continue to keep $\sigma$ fixed, and choose $a$ and $b$ to maximize (9). This is obtained by choosing $a$ and $b$ to minimize $\max_x|a+b'x-m_x|$. Therefore,

$$\sup_{h\in H}\inf_{\operatorname{supp} h}\frac{\tilde{g}}{h}=\frac{\sigma}{s}\exp\left[\frac{-\min_{a,b}\max_x(a+b'x-m_x)^2}{2(s^2-\sigma^2)}\right]. \tag{10}$$

That is, the minimum mixture estimates of the regression parameters are the minimax estimates for every fixed value of $\sigma$.

Finally, $\sigma$ should be selected to maximize (10). If one denotes $\sigma/s$ by $t$ and $\min_{a,b}\max_x(a+b'x-m_x)^2/2s^2$ by $u$, then the unique value of $t$ maximizing (10) is

$$t=\sqrt{1+u-\sqrt{(1+u)^2-1}}. \tag{11}$$

With this choice, $\sigma < s$ holds true.

For properties of the minimax, or Chebyshev, regression, see the references cited in the Introduction of this paper, and also Watson (1980). Because the minimax estimate is also the minimum mixture estimate, whenever minimax estimates are computed they always can be given minimum mixture interpretation and algorithms for minimax estimation can be used to compute minimum mixture estimates. Also, the mixture index of fit provides a natural measure of fit whenever minimax regression is used.

## 4 Minimum mixture estimation in regression with uniform error

In this section, similarly to the previous one, a simple regression model is considered but the error structure is assumed to be uniform and minimum mixture estimation is used to estimate the parameters.

The observations to analyze are, again, $(x, y)$ pairs, where $x$ is possibly vector valued and $y$ is considered to be a response variable to $x$. The family $F$ is defined as above and the primary family of interest, $G$, consists of distributions with the property that the conditional distribution of $Y$, given the value $X = x$, is uniform on an interval of length $2\delta$ for some $\delta > 0$ and the possible values of $X$ are $x_1, \ldots, x_k$. The value of the conditional density of $Y$ given $X = x$ for $x = x_1, \ldots, x_k$ and $y \in [\mu_x - \delta, \mu_x + \delta]$ is $1/2\delta$. The parameters of the conditional distribution are $\mu_x$ and $\delta$.

One convenient way to define joint distributions of $X$ and $Y$ with the same likelihood estimates as the ones obtained from the conditional densities, is to assign uniform marginal distribution to $X$. Therefore, the densities in $G$ have the following form:

$$g(x, y) = \frac{1}{k}\frac{1}{2\delta} \quad \text{if } x \in \{x_1, \ldots, x_k\} \quad \text{and} \quad y \in [\mu_x - \delta, \mu_x + \delta], \tag{12}$$

and zero otherwise.

The model, $H$, consists of distributions which, in addition to belonging to $G$, have the property that the conditional expectation of $Y$ is a linear function of the value of $X$, that is, $\mu_x = E(Y|X = x) = a + b'x$, with some unspecified constants $a$ and $b$, where $b$ is a vector of appropriate dimension.

That is, the primary family of interest is the family of distributions with uniform conditional distributions of $Y$ given the value of $X$, and the model of interest is the model of linear regression. Note that the assumption of equal lengths of the conditional ranges is parallel to the usual assumption in regression analysis with normal error structure that the conditional variance of $Y$ does not depend on the value of $X$.

To apply the mixture approach to measure the appropriateness of the assumption of linearity of the conditional expectations, just like in the previous section, a smoothed version of the data, that is, an estimate in $G$, $\tilde{g}$, is needed. When $\delta$ is known from theoretical considerations, previous experience or by design, only the $\mu_x$ parameters need to be estimated. A consistent estimate is the conditional mean of the observed $Y$ values, which is denoted by $m_x$. When $\delta$ is not known, its maximum likelihood estimate is the longest observed conditional range. The fact that this estimate maximizes the likelihood follows from the form (12) of the density. The sample will only have positive likelihood if the estimated conditional ranges cover all observed $y$ values, and the likelihood is maximal, if the common length of the estimated ranges is as short as possible. If for every observed $x$ value, there is only one observation for $Y$, an estimate of $\delta$ is not available. As it will be shown in the sequel, the minimum mixture estimates of the parameters $a$ and $b$ can still be determined in this case. Let $\tilde{\delta}$ denote either the fixed value of $\delta$, or its estimate, or an unspecified value, if an estimate is not available.

The density of the estimate $\tilde{g}$, has support

$$\mathop{\phantom{\bigcup}}_{i=1}^{k} \{x_i\} \times [m_{x_i} - \tilde{\delta}, m_{x_i} + \tilde{\delta}]$$

and in point $(x_i, y)$ of the support, the value of $\tilde{g}$ is

$$\frac{1}{2k\tilde{\delta}}.$$

A representation of the form (2) affords a $\pi$ value less than 1, only if

$$\operatorname{supp} h \quad \operatorname{supp} \tilde{g}. \tag{13}$$

According to the Theorem, to find the minimum mixture estimate $h^*$, and to determine the value of the mixture index of fit, the value of $\inf_{\mathrm{supp}\,h}(\tilde{g}/h)$ should be maximized in $h$ for $h \in H$. The value of the density for an $h \in H$ in any point of its support is $1/2k\varepsilon$, where $\varepsilon$ denotes the half length of the conditional ranges in $h$. If (13) holds true, then

$$\inf_{\mathrm{supp}\,h}(\tilde{g}/h) = \frac{1/(2\tilde{\delta})}{1/(2\varepsilon)} = \varepsilon/\tilde{\delta}.$$

For fixed $a$ and $b$, this is maximal if $\varepsilon$ is as large as possible, subject to (13). For fixed $a$ and $b$, (13) is fulfilled only if for $a + b'x \quad m_x$, $a + b'x - \varepsilon \quad m_x - \tilde{\delta}$, and for $a + b'x \quad m_x$, $a + b'x + \varepsilon \quad m_x + \tilde{\delta}$. That is, in order to fulfill (13), the following inequality should hold true:

$$\varepsilon \quad \tilde{\delta} - |a + b'x - m_x|,$$

in every point of the support of $h$, for fixed $a$ and $b$. Therefore, the possible maximum value of $\varepsilon$ is

$$\varepsilon_{\max} = \tilde{\delta} - \min_{a,b} \max_x |a + b'x - m_x|,$$

and therefore,

$$\sup_{a,b} \inf_{\mathrm{supp}\,h}(g/h) = \varepsilon_{\max}/\tilde{\delta} = 1 - \frac{\min_{a,b} \max_{x_i} |a + b'x_i - m_{x_i}|}{\tilde{\delta}}. \tag{14}$$

That is, the minimum mixture estimate of the regression parameters $a$ and $b$ is the minimax estimate. Note that this holds true whether or not the value of $\tilde{\delta}$ is known.

When $\tilde{\delta}$ is also known, as it will be the case in most applications, the Theorem and (14) imply that the value of the mixture index of fit is

$$\hat{\pi}^* = \frac{1}{\tilde{\delta}} \min_{a,b} \max_{x_i} |a + b'x_i - m_{x_i}|. \tag{15}$$

The estimated value of the mixture index of fit is proportional to $\tilde{\delta}$ and therefore, even if $\tilde{\delta}$ is not known, (15) can be used to assess relative importance in stepwise variable selection, as it will be illustrated in the next section.

Just like in the case of normal error, the minimax estimates of the regression parameters are the minimum mixture estimates.

## 5 Data analysis

The example presented in this section illustrates how the mixture index of fit can be used in minimax regression, including the problem of variable selection. Minimax regression is usually applied without distributional assumptions, and therefore little or no statistical inference is possible in these cases. As it is implied by the results of the two foregoing sections, whether the error structure is assumed to be normal or uniform, the minimax estimates optimize fit from the point of view of the mixture index of fit. This may provide the researcher with further insight into the appropriateness of the regression model.

The data of this example are taken from Wood (1973, Table B1). In this set of data, the dependent variable ($Y$) is the octane number of the product of a petrol refinery unit, and there are four independent variables, that is, $X$ is four dimensional. The independent variables describe various aspects of the production process. It was known that the designed range of output of the refinery unit is 90–94 octanes. Narula, Wellington (1985) applied a minimax estimation procedure of the regression parameters. They also considered a stepwise variable selection, which resulted in the following subsets of the variables: $\{X_1\}$, $\{X_1, X_4\}$, $\{X_1, X_2, X_4\}$ and $\{X_1, X_2, X_3, X_4\}$. For these sets of predictors, they obtained the following values for the minimal value of the maximum absolute error (MMAE): 1.7016, 0.9575, 0.9187 and 0.9149, respectively.

How good is the fit of the regression model based on all the four explanatory variables and how much worse is the fit of the regressions based on subsets of the variables only? How much is gained by including, say, a third explanatory variable in addition to two? These and similar questions cannot be easily answered based on MMAE.

It was shown in the previous sections, that the minimax estimates are also minimum mixture estimates, if normal or uniform error distribution is assumed. Under normality, if an estimate $s$ is available, the value of the mixture index of fit can be obtained by plugging in (11) into (10). Unfortunately, the value of $\sigma$ is not known for this set of data, and as there is only one observation for every value of $X$, no estimate $s$ can be computed. One may conclude, that under the assumption of normality, one does not have sufficient information to compute the value of the mixture index of fit and to base inference on its value. Furthermore, the assumption of normality, though customary with data which arose from physical processes, may be questioned on the basis on the very short support of the distribution of $Y$, namely 90–94 octanes. For a discussion on how the shape of the error distribution should influence the choice of estimation procedure, and further references, see Narula, Wellington (1985).

One possible choice for the error distribution with finite support is the uniform distribution. This assumption facilitates inferential procedures in this case. It is implied by (15) that $\hat{\pi}^* = \text{MMAE}/\tilde{\delta}$. The ratio of the estimated smallest fractions outside of two models measure relative gain in a readily interpretable sense. When $X_4$ is entered, that is, in the second step, the estimated smallest fraction which is outside of the model decreased by 44%(=(1.7016 – 0.9575)/1.7016). In the next step, the decrease is only 4% and in the last step is 0.4%. While it is somewhat difficult to judge whether a certain reduction in the value of MMAE warrants using one more explanatory variable in the regression equation, the mixture approach outlined in this paper yields the result, that the same amount of relative reduction takes place for the smallest possible fraction outside of the model as in the value of MMAE. Accordingly, entering $X_4$ appears to improve fit substantially, but entering the next two variables appears to result in little or no improvement.

*Acknowledgments:* The author is indebted to Ulf Böckenholt for helpful discussions and to the reviewers for making valuable comments on earlier versions of the manuscript. This research was supported in part by Grant No. OTKA T-016032 from the Hungarian National Science Foundation and was done while the author worked at the Center for Applied Statistics, Central European University .

## References

Clogg CC, Rudas T, Matthews S (1998) Analysis of contingency tables using graphical displays based on the mixture index of fit. In: Blasius J, Greenacre M (eds) Visualisation of Categorical Data, Academic Press, pp. 421–440

Clogg CC, Rudas T, Xi L (1995) A new index of structure for the analysis of mobility tables and other cross-classifications. In: Marsden P (ed) Sociological Methodology, Blackwell, Oxford, pp. 197–222

Fabian V (1988) Polynomial estimation of regression functions with the supremum norm. Annals of Statistics 16:1345–1368

Neter J, Kutner MH, Nachtsheim CJ, Wasserman W (1996) Applied Linear Statistical Models. Chicago, Irwin

Narula SC, Wellington JF (1985) A branch and bound procedure for selection of variables in minimax regression. SIAM J. Sci. Stat. Comput. 6:573–581

Rudas T, Clogg, CC, Lindsay, BG (1994) A new index of fit based on mixture methods for the analysis of contingency tables. Journal of the Royal Statistical Society Ser B 56:623–639

Rudas T, Zwick R (1997) Estimating the importance of differential item functioning. Journal of Educational and Behavioral Statistics 22:31–45

Schervish MJ (1996) P values: What they are and what they are not. The American Statistician 50:203–206

Sklar MG, Armstrong RD (1984) An algorithm for discrete Chebyshev curve fitting for the simple model using a dual linear programming approach. Communications in Statistics Part B 13:555–569

Wood FS (1973) The use of individual effects and residuals in fitting equations to data. Technometrics 15:677–695

Xi L (1996) Measuring goodness-of-fit in the analysis of contingency tables with mixture-based indices: algorithms, asymptotics and inference. Ph. D. Thesis, Department of Statistics, The Pennsylvania State University

Watson, GA (1980) Approximation Theory and Numerical Methods. Wiley, New York